The inefficacy of superficial similarities for improving instructor-student relationships*

Replicating 'Taking It to the Next Level'

Amy Farrow

April 13th, 2021

Abstract

This paper replicates the 2019 article 'Taking It to the Next Level' (Robinson, Scott, and Gottfried 2019), which evaluates an intervention to improve college instructor-student relationship. The modeling results indicate that the intervention, consisting of informing instructors and students about commonalities, has a weak positive effect on student perceptions of instructor-student similarity, but no effect on student perceptions of instructor-student relationship, instructor perception of similarity or instructor-student relationship, grades, or re-enrollment. While the scalability and affordability of the intervention are desirable, there are no results in any of the targeted measures: those that affect and reflect college retention. These results are consistent with the original paper. TODO: revise to clearliy indicate my own findings

Keywords: instructor-student relationship, college, replication study

Contents

*Code and data are available at: github.com/amycfarrow/takingittothenextlevelrepro.

1 Introduction

College is perceived to be a meritocratic tool for social mobility and career success (Bowen, Chingos, and McPherson 2009). Unfortunately, retention is a large problem in contemporary American colleges, and many students begin degrees without completing them; six-year completion rates for full-time first-time students range from 51% to 86%, depending on school (Delbanco 2015). Even when controlling for pre-college test scores and initial enrollments, completion rate disparities exist based on parental education, socio-economic status, and race/ethnicity (Delbanco 2015). Disparities in college completion lead to further entrenchment of long-standing inequalities (Delbanco 2015). Thus, measures to help students persist in college completion are desirable to reduce wasted resources and to increase societal equity.

Robinson, Scott, and Gottfried (2019)'s paper, "Taking It to the Next Level: A Field Experiment to Improve Instructor-Student Relationships in College", tests an intervention to improve college retention and performance. In this field experiment, they tested the effect of instructor-student similarity on instructor-student relationships (ISRs) and measures of student success. Based on extensive K-12 research about the importance of instructor-student relationship for student success, Robinson, Scott, and Gottfried (2019) aimed to establish how instructor-student relationships could be improved at the college level, and to test if this improvement had a positive result. The experiment consisted of a randomized controlled trial where some undergraduate students were informed of similarities they shared with their instructor, while others were not. Student and instructor perceptions of similarity and the instructor-student relationship were measured through surveys, and student performance measures were collected from school records.

This paper replicates Robinson, Scott, and Gottfried (2019)'s original analysis, using anonymized data provided by the authors. First, methods are discussed, and the measures and scales created by Robinson, Scott, and Gottfried (2019) are evaluated. Second, different demographics of instructor and student populations are explored, similarity of treatment and control groups is established, outcomes of interest are compared for treatment and control groups, and missing data is evaluated. Third, Robinson, Scott, and Gottfried (2019)'s models are replicated, and additional models are explored. Linear, logistic, and ordinal logistic models are replicated to predict outcomes including perceived ISR, grade, and persistence. These models show that treatment slightly improves student perception of instructor-student similarity, but does not significantly affect student or instructor ISR perception, grade, exam grade, or future enrollment. Replicating exploratory models, the relationship between similarity and relationship, as well as between relationship and student outcomes, are modeled. Additional models use different controls; specifically, variables indicating shared racial and gender self-identification. As with the replication models, treatment has no significant impact on ISR perception or student outcomes.

1.1 Literature review

TODO: expand this paragraph about interventions, preferably targeting similar ideas to the Robinson, Scott, and Gottfried (2019) paper. What has worked? Draw attention to results. Make sure to refer to any studies that Robinson, Scott, and Gottfried (2019) say are especially influential for their work.

Many interventions to increase completion rates have been suggested and tested. For example, Evans et al. (2020) studied case managment techniques; Gurantz, Hurwitz, and Smith (2017) investigated the effect of high school recruitment; Bers and Schuetz (2014) identify the importance of interpersonal relationships; Gilbert and Horn (2020) argue that certificates awarded prior to degrees can ameliorate the 'college-

completion crisis'. These are only a small sample of the interventions suggested. Implementing an effective program is difficult due to complicated causes of attrition, embedded social inequalities, and expenses.

TODO: write a paragraph about research on why similarity builds relationships (similarity -> relationships)

TODO: write a paragraph about research on why relationships matter for scholastic performance (relationships -> grades) Abrami and Mizener (1985)'s study of college students found a . One study of Taiwanese high school students explored the link between shared education and life values and student performance (Lai 2015). While they found a significant, but small, positive relationship between life values and student performance on analytic tests, education values gave mixed results depending on the natures of the education value and the analytic test in question (Lai 2015).

TODO: write a paragraph about research on why relationships matter for persistence (relationships -> persistence)

2 Methodology

Robinson, Scott, and Gottfried (2019) conducted a randomized controlled trial to assess the impact of awareness of instructor-student similarities on perceived similarity, instructor-student relationship, course grade, and re-enrollment.

2.1 Participants

The study took place in the 2017 spring semester at a large Californian University. The study included 120 instructors and their 2,749 students. The instructors participated in the study based on interest and a gift-card incentive, and their students were invited to participate unincentivized. Students were only enrolled in the study for one class, in the event that they were taking classes with multiple participating instructors. The initial sample of 147 instructors was a convenience sample: the study was advertised, interested instructors signed up and were enrolled if their course met the study requirements. 145 instructors consented and took the initial survey, resulting in an corresponding sample of 3,352 students. Once errors of administration, missing responses, and inadequate time spent on surveys were considered, the sample consisted of only 119 instructors and 2,273 students.

2.2 Treatment and control

Participating students were randomly assigned to either treatment or control. At the beginning of the term, all participating students and instructors were given "get to know you" surveys. Using those responses, for each student in the treatment group, seven commonalities were identified between student and instructor (for example, perhaps both student and instructor binge-watch TV to relieve stress, or appreciate loyalty as the most important friend quality), and both student and instructor were informed of these commonalities. They completed a few questions about the similarities and were reminded of them through the semester to ensure they were internalized. Students in the control group were informed about similarities they shared with students in another part of the country, and instructors were told nothing about these students.

2.3 Procedures

All students participated in a survey immediately following the treatment or the placebo. They were surveyed again at the end of the course. Instructors were surveyed only at the end of the course.

2.4 Measures

Robinson, Scott, and Gottfried (2019) identify key measures. Full descriptions of all measures are available in Appendix A.

Some are extracted from the student survey at the beginning of the term:

1. Immediately after the treatment or placebo, students answered six questions about their perceived similarity to the instructor, on scales of 1 to 5. These responses were averaged to create a student similarity perception scale.
2. Immediately after the treatment or placebo, students answered seven questions about their perception of the instructor-student relationship, on scales of 1 to 5. These responses were averaged to create a student ISR perception scale.
3. Student gender

Others are extracted from the student survey at the end of the term:

4. At the end of the semester, students answered the student similarity perception scale questions again.
5. At the end of the semester, students answered seven questions about their perception of the instructor-student relationship, on scales of 1 to 5. These responses were averaged to create a student ISR perception scale.

Others come from the instructor survey at the end of the term:

6. Instructor similarity perception: At the end of the semester, instructors answered only one question about similarity with the student, on a scale of 1 to 5.
   - Overall, how similar do you think you and STUDENTNAME are?
7. Instructor ISR perception scale: At the end of the semester, instructors answered seven questions about their perception of the instructor-student relationship, on scales of 1 to 5. These responses were averaged to create an ISR scale.
8. Final grade: Instructors were asked to report the student's grade on their final exam, paper, or project.

Finally, some are extracted from the university's internal records:

9. The final grade that the student received in the course.
10. The student's final grade, standardized against other grades in the course.
11. The student's cumulative GPA (CGPA) after the Fall 2016 term.
12. Persistence: The student's status as of Fall term 2017: not enrolled or enrolled.

2.4.1 Scale Reliability

A scale is externally reliable if it gives the same results across different testings (Coolican 2014). Test-retest reliability, where a scale is given to a group at different times, is a common way to assess this (Coolican 2014). In our data, there is a retest, but interactions between instructors and students may reasonably be expected to affect the second testing, rendering the data unuseful for assessing test-retest reliability.

A scale is internally reliable if it is consistent within itself (Coolican 2014). Cronbach's alpha was used to assess the reliability of the similarity and ISR scales. This statistic evaluates how much participants vary on individual items, compared to how they vary overall (Coolican 2014). Good reliability is indicatedby an alpha between .75 and 1 (Coolican 2014). For the initial student similarity scale, initial student ISR scale, end-of-term student similarity scale, end-of-term student ISR scale, and end-of-term student ISR scale, the alpha values were 0.892, 0.872, 0.91, 0.905, and 0.926, respectively, indicating that the scales are reliable.

2.4.2 Scale Validity

A scale is valid if it measures what it is supposed to measure (Coolican 2014). In this case, the scales should measure perceived similarity and perceived instructor-student relationship (ISR). Robinson, Scott, and Gottfried (2019) used modified versions of the scales from Gehlbach et al. (2015), where they were used with ninth-grade American students and their teachers; the scales were originally were presented in Gehlbach, Brinkworth, and Harris (2012), where they were developed for use with sixth-, seventh-, and eight-grade American students and their teachers. Given the fact that these scales were developed for children, it is worth considering what the scales actually measure. To do this, we can look at the specific questions in the three distinct scales.

To assess similarity, students were asked about values, course goals, views on course content, general commonalities, personality, and overall similarity. By this standard, high similarity is when a student shares

values, course goals, and views on course content, and the student and instructor have similar personalities, high commonality, and high general similarity. Questions ask about course-specific measures (content and goals), personal traits (values and personality), and non-specific traits (commonalities and general similarity), but do not ask about life experience, abilities, approaches, amoungst other ways that two people can theoretically be similar.

TODO: ˆ link to any other definitions/studies of similarity

To assess ISR, students were asked about enjoyment of learning, friendliness, encouragement, excitement, motivation, caring, and overall learning. By this standard, a high-quality ISR is one where a student enjoys learning from, is motivated by, and learns a great deal from a friendly, encouraging, caring instructor, who would be excited to see them in three years time. Questions do not ask about respect, reliability, admiration, challenge, personal growth, frequency of contact, or many other aspects of how a student perceives an instructor-student relationship, especially at the college level.

To assess ISR, instructors were asked about enjoyment of aiding learning, caring, frequency of encouragement, friendliness, excitement, motivation, and overall learning. By this standard, a high-quality ISR is one where an instructor enjoys helping and frequently encourages a friendly, caring student who is motivated and learns a great deal from this instructor, and the instructor would be excited to see the student in three years. Questions do not ask about respectfulness, interest, potential, frequency of contact, tone of that contact, or other aspects of instructor perception of an instructor-student relationship.

TODO: ˆˆ link to any other definitions of ISR

3   Data

TODO: add all citations for R packages, R, and R markdown The correlation matrix was made using functions by Engineering (2018).

The data, in a .dta format, and Stata code for the Robinson, Scott, and Gottfried (2019) paper are available on the Inter-University Consortium for Political and Social Research (Carly D. Robinson, Scott, and Gottfried 2020). The dataset contains 36838 observations of 653 variables.

3.1   Demographics

Table 1 shows the student covariates for the treatment and control samples. There are no significant differences between the treatment and control groups.

Table 2 shows the instructor covariates. Because instructors are counted multiple times, for each student in the treatment and control groups that is in their course, only the totals are shown. The same covariates, presented by treatment and control groups, are shown in Appendix B.

There are differences between the students and instructors. Notably, the student sample is 62.7% female, while the instructor sample is 78.2% female. The student sample is 19.8% White, while the instructor sample is 38.7% White. The student sample is 47% Hispanic, while the instructor sample is only 3.4%. Finally, the student sample is 9.8% multiracial, but the instructor sample is 52.1% multiracial. Some of these differences may be due to self-identification: while one person who is Hispanic may select only Hispanic, another may select, for example, Hispanic as well as White, and be classed differently as a result. For this reason, it is very difficult to draw conclusions. Additionally, students are 43.5% first-generation college students, while only 25.2% of instructors are.

In Figure 1, we can see that more students identified as Hispanic, while more instructors identified as White and multi-racial.

In Figure 2, we can see that most class sizes are between 25 and 50 students, but there are some that are much larger.

Table 1: Student covariates for treatment and control groups

Group	N	Missing		Control	Treatment	Total	p
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273	
Student gender	2273	0	Male	432 (38.0)	416 (36.6)	848 (37.3)	0.526
			Female	705 (62.0)	720 (63.4)	1425 (62.7)	
Student race	2273	0	American Indian	1 (0.1)		1 (0.0)	0.669
			Asian	126 (11.1)	124 (10.9)	250 (11.0)	
			Black	51 (4.5)	43 (3.8)	94 (4.1)	
			Black and White	4 (0.4)	6 (0.5)	10 (0.4)	
			Black Hispanic	4 (0.4)	4 (0.4)	8 (0.4)	
			Hispanic	513 (45.1)	556 (48.9)	1069 (47.0)	
			Middle Eastern	54 (4.7)	55 (4.8)	109 (4.8)	
			Other	40 (3.5)	31 (2.7)	71 (3.1)	
			Other multi	78 (6.9)	66 (5.8)	144 (6.3)	
			Unknown	2 (0.2)	3 (0.3)	5 (0.2)	
			White	237 (20.8)	213 (18.8)	450 (19.8)	
			White Hispanic	27 (2.4)	35 (3.1)	62 (2.7)	
Student first-gen status	2270	3	No	634 (55.8)	648 (57.1)	1282 (56.5)	0.550
			Yes	502 (44.2)	486 (42.9)	988 (43.5)	
Year	2246	27	1	78 (6.9)	74 (6.6)	152 (6.8)	0.977
			2	178 (15.8)	180 (16.1)	358 (15.9)	
			3	191 (17.0)	182 (16.2)	373 (16.6)	
			4	665 (59.1)	668 (59.6)	1333 (59.3)	
			5	14 (1.2)	16 (1.4)	30 (1.3)	
Student age	2269	4	Mean (SD)	22.6 (5.4)	22.7 (5.3)	22.7 (5.4)	0.726
CPGA	2178	95	Mean (SD)	3.0 (0.6)	3.0 (0.7)	3.0 (0.6)	0.414

Table 2: Teacher covariates

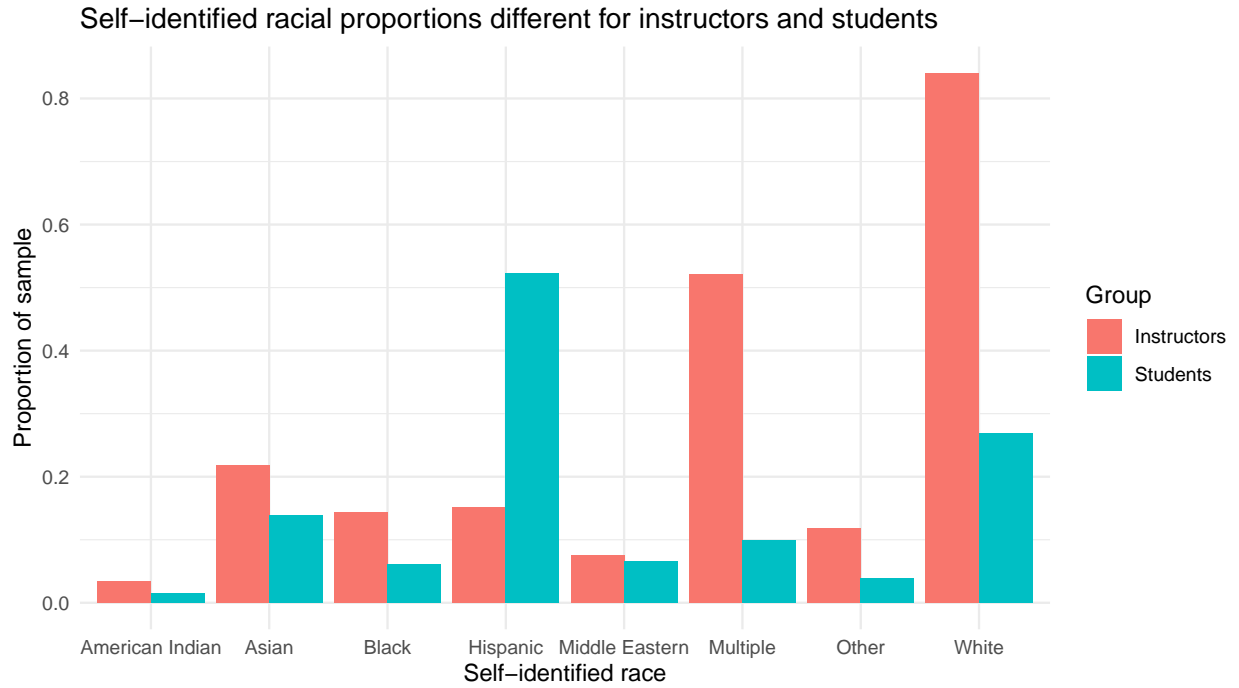|	N	Missing		
				N = 119
Instructor gender	119	0	Male	26 (21.8)
			Female	93 (78.2)
Instructor race	119	0	Asian	4 (3.4)
			Black	2 (1.7)
			Black and White	9 (7.6)
			Black Hispanic	1 (0.8)
			Hispanic	4 (3.4)
			Other multi	44 (37.0)
			White	46 (38.7)
			White Hispanic	8 (6.7)
			Unknown	1 (0.8)
Instructor first-gen status	119	0	No	89 (74.8)
			Yes	30 (25.2)
Instructor age	118	1	Mean (SD)	44.5 (11.0)
Course size	118	1	Mean (SD)	36.4 (27.9)

Figure 1: Bar chart illustrating racial identification proprtions of participants

Figure 2: Histogram illustrating class sizes

Table 3: Outcomes of interest for treatment and control groups

	N	Missing		Control	Treatment	Total
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273
Initial student perception						
Similarity	2272	1	Mean (SD)	3.4 (0.7)	3.6 (0.7)	3.5 (0.7)
ISR	2273	0	Mean (SD)	4.1 (0.6)	4.1 (0.6)	4.1 (0.6)
End-of-term student perception						
Similarity	1831	442	Mean (SD)	3.5 (0.8)	3.6 (0.8)	3.6 (0.8)
ISR	1831	442	Mean (SD)	4.1 (0.7)	4.1 (0.7)	4.1 (0.7)
End-of-term instructor perception						
Similarity	2225	48	1	133 (11.9)	132 (11.9)	265 (11.9)
			2	344 (30.8)	343 (31.0)	687 (30.9)
			3	402 (36.0)	400 (36.1)	802 (36.0)
			4	171 (15.3)	166 (15.0)	337 (15.1)
			5	67 (6.0)	67 (6.0)	134 (6.0)
ISR	2240	33	Mean (SD)	3.5 (0.9)	3.5 (0.9)	3.5 (0.9)
Student outcomes						
Course grade	2147	126	Mean (SD)	3.1 (1.0)	3.1 (1.0)	3.1 (1.0)
Stand. grade	2120	153	Mean (SD)	0.1 (0.9)	0.1 (0.9)	0.1 (0.9)
Final grade	2109	164	Mean (SD)	3.0 (1.1)	3.0 (1.1)	3.0 (1.1)
Peristence	1786	487	No	59 (6.7)	68 (7.5)	127 (7.1)
			Yes	823 (93.3)	836 (92.5)	1659 (92.9)

3.2 Outcomes of interest

Table 3 shows summary statistics for key variables identified by Robinson, Scott, and Gottfried (2019), including missing values.

Figure 3 shows key continuous outcome distributions for both treatment and control. The only visible differences are in the initial student similarity perception and end-of-term student similarity perception.

Replicating Robinson, Scott, and Gottfried (2019)'s study, Table 4 displays a correlation matrix for the measures that Robinson, Scott, and Gottfried (2019) identified as significant. Some measures, unsurprisingly, are highly correlated: grade and standardized grade, and grade and final exam grade. Others are moderately correlated: student initial perceptions of similarity and instructor-student relationship, student end-of-term perceptions of similarity and instructor-student relationship, and instructor end-of-term perceptions of similarity and instructor-student relationship. These last three pairs could indicate that the similarity perception scales and instructor-student relationship perception scales may not be measuring distinct concepts, or that feelings of similarity and positivity in relationships are strongly associated in the classroom.

3.3 Missing data

There are no significant patterns in missing data. Appendix C shows the distributions of key variables relative to missing data in other key variables.

3.4 Attrition

129 eligible instructors completed the first survey, but 6 of those instructors did not complete the second survey. Corresponding to the remaining 123 instructors, 2,801 students completed the first survey (1,392 in the treatment group, 1,409 in the control group), but 682 of those students did not complete the second survey (328 in the treatment group, 354 in the control group). The attrition in the treatment and control groups was comparable (24% in the treatment group, 25% in the control group).
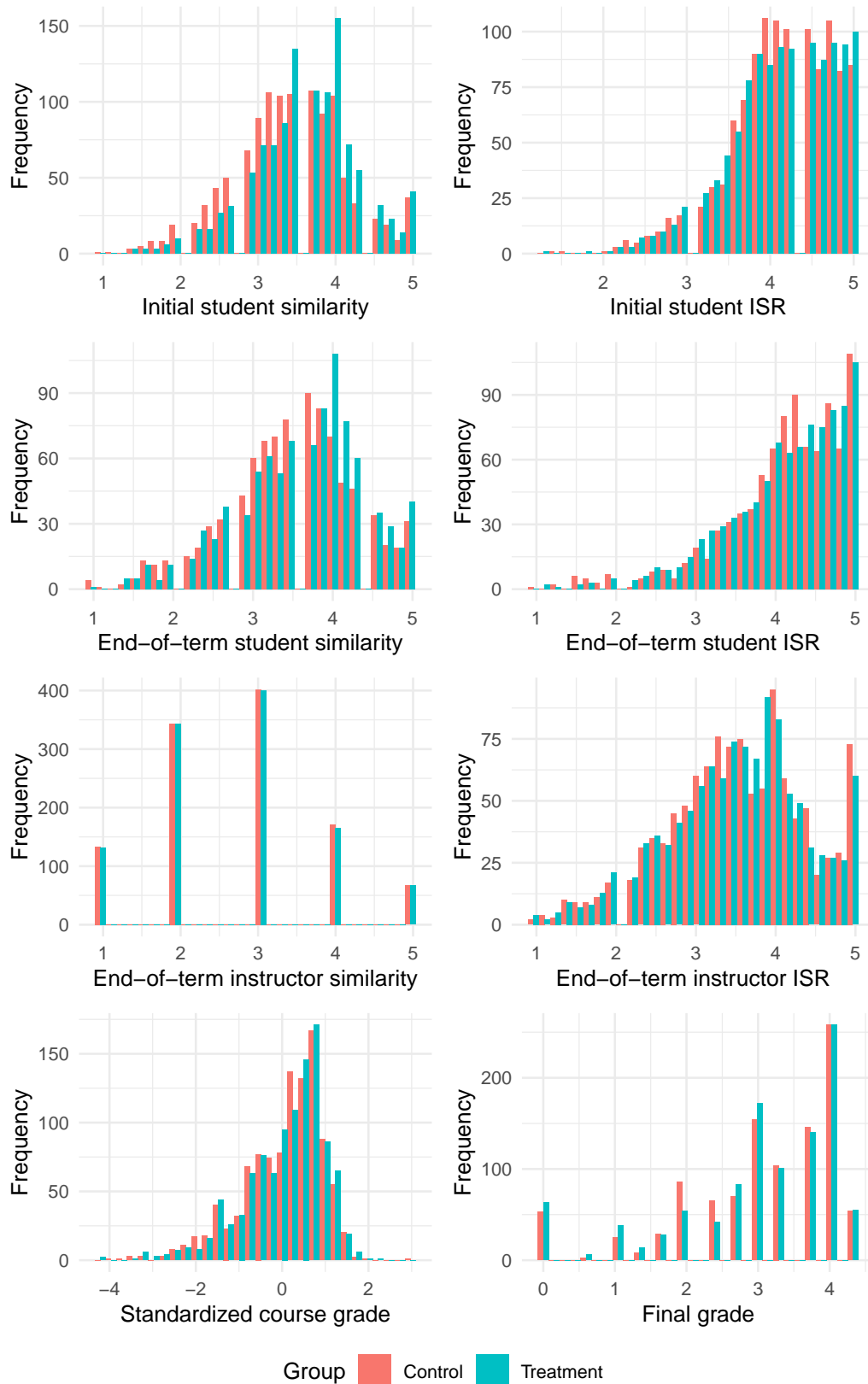
Figure 3: Histograms illustrating continous and ordinal key outcomes

Table 4: Correlation matrix for continuous and ordinal key variables and outcomes

	N	Missing	Mean (SD)	Initial student		End-of-term student		End-of-term instructor		Grades				Course
				Similarity	ISR	Similarity	ISR	Similarity	ISR	Course	Stand. course	Final	CGPA	Size
Initial student perception														
Similarity	2272	1	3.5 (0.7)	-										
ISR	2273	0	4.1 (0.6)	0.67***	-									
End-of-term student perception														
Similarity	1831	442	3.6 (0.8)	0.58***	0.5***	-								
ISR	1831	442	4.1 (0.7)	0.43***	0.57***	0.79***	-							
End-of-term instructor perception														
Similarity	2225	48	2.7 (1.0)	0.15***	0.16***	0.25***	0.24***	-						
ISR	2240	33	3.5 (0.9)	0.19***	0.21***	0.29***	0.33***	0.68***	-					
Grades														
Course	2147	126	3.1 (1.0)	0.07**	0.1***	0.2***	0.23***	0.38***	0.44***	-				
Stand. course	2120	153	0.1 (0.9)	0.02	0.01	0.12***	0.11***	0.29***	0.35***	0.81***	-			
Final	2109	164	3.0 (1.1)	0.06**	0.1***	0.17***	0.18***	0.38***	0.43***	0.81***	0.62***	-		
CGPA	2178	95	3.0 (0.6)	0.05*	0.03	0.08***	0.06*	0.19***	0.19***	0.43***	0.41***	0.39***	-	
Course														
Size	2269	4	49.5 (44.6)	0.02	0.01	0.04	0.01	0.19***	0.11***	-0.01	0	- 0.05*	-0.02	-

3.5 Data selection

From the 36,838 observations, or potential units of study (corresponding to all undergraduate student records of the university), only 2,273 were used in analysis. Units of study were excluded because the instructor did not participate in the study (33,486); the student did not consent (123); the student did not spend more than a second reading the consent page (1); the initial student survey, initial instructor survey, or end-of-term instructor survey was not complete (428); the instructor mistakenly administered the wrong survey (30); the course was online (20); the course was for graduate students (50); or a participant did not spend more than ten seconds reading and answering a page with five or more questions (427). The ten-second time limit is somewhat arbitrary, and a different choice could lead to different study results.

4 Models

4.1 Replication models

$treatment_i$ is the indicator that treatment was given.

X_{1i} is a vector of student-level covariates (anticipated student ISR, gender, CGPA). Anticipated ISR is used as a control for end-of-term student ISR perception, while gender and CGPA are used as controls for grade based outcomes. Robinson, Scott, and Gottfried (2019) state this is because females earn higher grades generally.

X_{2j} is a vector of instructor-level covariates (course size, teacher ID).

ϵ_{ij} is a clustered residual.

β_0, $beta_1$, Γ_1, Γ_2, and a_k are coefficients on the resulting models.

4.1.1 Linear models

Equation (1) is Robinson, Scott, and Gottfried (2019)'s linear model, used for continuous outcomes. This include complete scale outcomes: immediate student similarity rating (s1_sim), end of semester student similarity rating (s2_sim), end of semester student ISR rating (s2_tsr), and end of semester instructor ISR rating (t2_tsr). Because these scales are created using 6-7 questions answered on a scale of 1 to 5, they have a total of 25 or 30 possible values, they can be treated as continuous variables and modeled using linear functions. It also includes grade-based outcomes: course grade (grade) and objectively graded exam grade (t2_finalexam), both of which are shown on a 4.0 GPA scale.

$$Outcome_{ij} = \beta_0 + \beta_1 treatment_i + X_{1i}\Gamma_1 + X_{2j}\Gamma_2 + \epsilon_{ij} \qquad (1)$$

4.1.2 Ordinal logistic models

Equation (2) is Robinson, Scott, and Gottfried (2019)'s ordinal logistic model, used for the ordinal outcome, which is end of semester instructor similarity rating (t2_sim1).

$$prob(outcome_{ij}) = a_k + \beta_1 treatment_i + X_{1i}\Gamma_1 + X_{2j}\Gamma_2 + \epsilon_{ij} > k \qquad (2)$$

4.1.3 Logistic models

Equation (3) is Robinson, Scott, and Gottfried (2019)'s logistic model, used for the binary outcome, which is enrollment in Fall term 2017 (f17_enrolled).

$$prob(outcome_{ij}) = a_k + \beta_1 treatment_i + X_{1i}\Gamma_1 + X_{2j}\Gamma_2 + \epsilon_{ij} > k \qquad (3)$$

4.2 Additional models

When creating additional models, I focused on three key outcomes: end-of-term student similarity perception scale, end-of-term student ISR perception scale, and standardized course grade. I selected these because interventions like Robinson, Scott, and Gottfried (2019)'s are aimed at improving student outcomes (like course performance) and perceptions that could improve student outcomes (like student ISR perception), and this particular intervention tries to accomplish these goals by improving perceived similarity.

$$Outcome_{ij} = \beta_0 + \beta_1 treatment_i + X_{1i}\Gamma_1 + X_{2j}\Gamma_2 + X_{3jk}\Gamma_3 + \epsilon_{ij} \qquad (4)$$

Equation (4) shows the linear model used. X_{3jk} represents student-instructor commonality covariates (matching racial self-ID, matching gender self-ID, age difference).

5 Results

5.1 Replication results

In Table 5, Model 1 shows a statistically significant relationship between treatment and student initial similarity perception. However, the expected increase is only 0.158 on a scale of 1 to 5. Model 2 shows a statistically significant relationship between treatment and student end-of-semester similarity perception. However, the expected increase is only 0.101 on a scale of 1 to 5.

Model 3 shows no significant relationship between treatment and student end-of-semester ISR perception. However, it does show a significant correlation between anticipated student ISR and student end-of-semester ISR. For every 1 point increase in anticipated ISR on a 1 to 5 scale, the expected increase in end-of-semester ISR is 0.674.

Models 4, 5, 6, 7, and 8 show no significant relationship between treatment and instructor end-of-semester ISR perception, course grade, objectively graded exam grade, instructor perception of similarity at the end of the term, and enrollment in the subsequent semester.

Table 6 considers the relationship between perceived similarity and perceived ISR and the relationship between perceived ISR and student outcomes. Because only the treatment was randomized, we cannot make causal inferences.

In Models 1 and 2, we can see that student and instructor perceived similarity are both significantly related to student and instructor perceived ISR. Unsurprisingly, student perceived similarity is a better predictor of

Table 5: Replication model results

	Dependent variable:							
	Student sim. 1	Student sim. 2	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Instructor sim. 2	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*ordered logistic*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.190***	0.118***	0.006	−0.019	0.043	0.005	−0.017	0.127
	(0.028)	(0.034)	(0.026)	(0.030)	(0.035)	(0.084)	(0.081)	(0.232)
Student ISR 1			0.565***					
			(0.024)					
Student female					−0.046	−0.146		−0.141
					(0.041)	(0.095)		(0.276)
CGPA					0.651***	0.603***		1.018***
					(0.029)	(0.062)		(0.162)
Course size	0.145**	0.158**	−0.131**	−0.039	0.055	0.010***	0.205***	−6.099
	(0.061)	(0.067)	(0.052)	(0.063)	(0.074)	(0.003)	(0.002)	(1,142.578)
Constant	−0.941	−1.694	4.771***	4.313**	−0.407	−0.483		182.465
	(1.739)	(1.920)	(1.469)	(1.802)	(2.098)	(0.396)		(34,277.330)
Teacher fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,268	1,829	1,829	2,236	2,063	514	2,221	1,687
R^2	0.141	0.237	0.495	0.391	0.400	0.415		
Adjusted R^2	0.094	0.185	0.460	0.357	0.364	0.391		
Log Likelihood								−277.677
Akaike Inf. Crit.								793.353

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Replication model results 2

	Dependent variable:				
	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.076***	−0.038	0.035	−0.036	0.306
	(0.020)	(0.024)	(0.032)	(0.085)	(0.308)
Student sim. 2	0.660***	0.119***			
	(0.014)	(0.018)			
Instructor sim. 2	0.034***	0.502***			
	(0.013)	(0.016)			
Course size	−0.102***	−0.052			
	(0.039)	(0.048)			
Student ISR 2			0.076***	0.144**	−0.018
			(0.027)	(0.065)	(0.250)
Instructor ISR 2			0.395***	0.432***	0.406
			(0.026)	(0.068)	(0.256)
Student female			0.018	−0.098	0.242
			(0.037)	(0.095)	(0.365)
CGPA			0.474***	0.407***	0.840***
			(0.027)	(0.063)	(0.218)
Constant	4.259***	2.941**	0.117	−1.366***	16.702
	(1.108)	(1.359)	(0.181)	(0.417)	(3, 560.318)
Teacher fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	1,809	1,809	1,665	394	1,343
R^2	0.713	0.653	0.491	0.486	
Adjusted R^2	0.693	0.629	0.453	0.456	
Log Likelihood					-159.172
Akaike Inf. Crit.					558.344

Note: *p<0.1; **p<0.05; ***p<0.01

student perceived ISR, and instructor perceived similarity is a better predictor of instructor perceived ISR. Additionally, course size is negatively related to student ISR perception.

Models 3 and 4 show that student ISR perception and instructor ISR perception are both significantly related to course grade and final grade. Instructor perception is a stronger predictor than student perception.

Model 5 indicates that ISR perception is not significantly related to persistence.

TODO: clustered residuals (low priority)

5.2 Additional results

Table 7: Additional model results

	Dependent variable:		
	Student sim. 2	Student ISR 2	Std. course grade
	(1)	(2)	(3)
Treatment	0.107***	−0.005	0.032
	(0.037)	(0.035)	(0.039)
CGPA			0.651***
			(0.030)
Course size	0.0004	−0.0001	−0.0004
	(0.0004)	(0.0004)	(0.0005)
Matching racial self-ID	−0.015	−0.099*	0.165**
	(0.060)	(0.056)	(0.064)
Matching gender self-ID	0.069*	0.053	0.110***
	(0.038)	(0.036)	(0.041)
Age difference	−0.007***	−0.006***	−0.008***
	(0.002)	(0.001)	(0.002)
Constant	3.586***	4.242***	1.220***
	(0.052)	(0.049)	(0.109)
Observations	1,819	1,819	2,065
R^2	0.020	0.014	0.200
Adjusted R^2	0.017	0.011	0.198
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7 shows the results of models that consider matching student-instructor traits. As with the replication results, there is a statistically significant relationship between treatment and student initial similarity perception. However, the expected increase is only 0.108 on a scale of 1 to 5, which is even lower than in the replication models. There is still no significant relationship between treatment and student end-of-semester ISR perception. Additionally, there is no significant relationship between treatment and standardized course grade.

TODO: Shiny.

6 Discussion

TODO: why did Robinson, Scott, and Gottfried (2019) pick their models? What do I think about the selection?

TODO: does this study have external validity?

TODO: Inefficacy of weak measures in the face of divergent student/instructor demographics and massive class sizes.

TODO: Who perceives similarity

TODO: What does it mean for relationships and therefore academic success to be built on similarity.

A   Appendix A

Complete list of key measures

- Initial student survey:
  1. Student similarity perception scale (s1_sim)
     - Overall, how similar to your instructor's values do you think your values are?
     - How similar are your goals for the course and your instructor's goals?
     - In general, how similar do you think your views about the course content and your instructor's are?
     - How much do you think you have in common with your instructor?
     - How similar do you think your personality is compared to your instructor's?
     - Overall, how similar do you think you and your instructor are?
  2. Student ISR perception scale (s1_tsr)
     - How much do you think you will enjoy learning from this instructor?
     - How friendly do you think this instructor will be towards you?
     - How encouraging do you think this instructor will be towards you?
     - If you came back to visit this instructor three years from now, how excited do you think they would be?
     - How motivating do you think you will find this instructor's class?
     - How caring do you think this instructor will be towards you?
     - Overall, how much do you think you will learn from this instructor?
  3. Student gender (s1_female)
- End of term student survey:
  4. Student similarity perception scale (s2_sim)
  5. Student ISR perception scale (s2_tsr)
     - How much do you enjoy learning from this professor?
     - How friendly do you think this professor is towards you?
     - If you came back to visit this professor three years from now, how excited do you think they would be?
     - How motivating do you find this professor's class?
     - How caring do you think this professor is towards you?
     - How encouraging do you think this professor is towards you?
     - Overall, how much do you think you have learned from this professor?
- End of term instructor survey:
  6. Instructor similarity perception (t2_sim1):
     - Overall, how similar do you think you and STUDENTNAME are?
  7. Instructor ISR perception scale (t2_tsr):
     - How much did you enjoy helping STUDENTNAME learn?
     - How caring was STUDENTNAME towards you?
     - How often did you say something encouraging to STUDENTNAME?
     - How friendly was STUDENTNAME towards you?
     - If this student came back to visit you three years from now, how excited would you be?
     - How motivating did STUDENTNAME find the activities that you plan for class?
     - Overall, how much did STUDENTNAME learn from you?
  8. Final grade (t2_finalexam): Instructors were asked to report the student's grade on their final exam, paper, or project.
- University internal records:
  9. Course grade (grade): The final grade that the student received in the course.
  10. Standardized course grade (std_grade): The student's final grade, standardized against other grades in the course.
  11. CGPA (ir_f16_gpa): The student's cumulative GPA after the Fall 2016 term.
  12. Persistence (f17_enrolled): The student's status as of Fall term 2017: not enrolled or enrolled.

B   Appendix B

Table 8: Teacher covariates for treatment and control groups

Group	N	Missing		Control	Treatment	Total	p
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273	
Teacher gender	2273	0	Male	230 (20.2)	229 (20.2)	459 (20.2)	1.000
			Female	907 (79.8)	907 (79.8)	1814 (79.8)	
Teacher race	2273	0	Asian	37 (3.3)	35 (3.1)	72 (3.2)	0.996
			Black	7 (0.6)	5 (0.4)	12 (0.5)	
			Black and White	96 (8.4)	97 (8.5)	193 (8.5)	
			Black Hispanic	12 (1.1)	10 (0.9)	22 (1.0)	
			Hispanic	35 (3.1)	34 (3.0)	69 (3.0)	
			Other multi	441 (38.8)	454 (40.0)	895 (39.4)	
			Unknown	24 (2.1)	21 (1.8)	45 (2.0)	
			White	419 (36.9)	420 (37.0)	839 (36.9)	
			White Hispanic	66 (5.8)	60 (5.3)	126 (5.5)	
Teacher first-gen status	2273	0	No	891 (78.4)	890 (78.3)	1781 (78.4)	1.000
			Yes	246 (21.6)	246 (21.7)	492 (21.6)	
Teacher age	2263	10	Mean (SD)	43.4 (10.5)	43.4 (10.6)	43.4 (10.5)	0.958
Course size	2269	4	Mean (SD)	49.3 (44.3)	49.7 (44.9)	49.5 (44.6)	0.812

C Appendix C

Missing data matrix
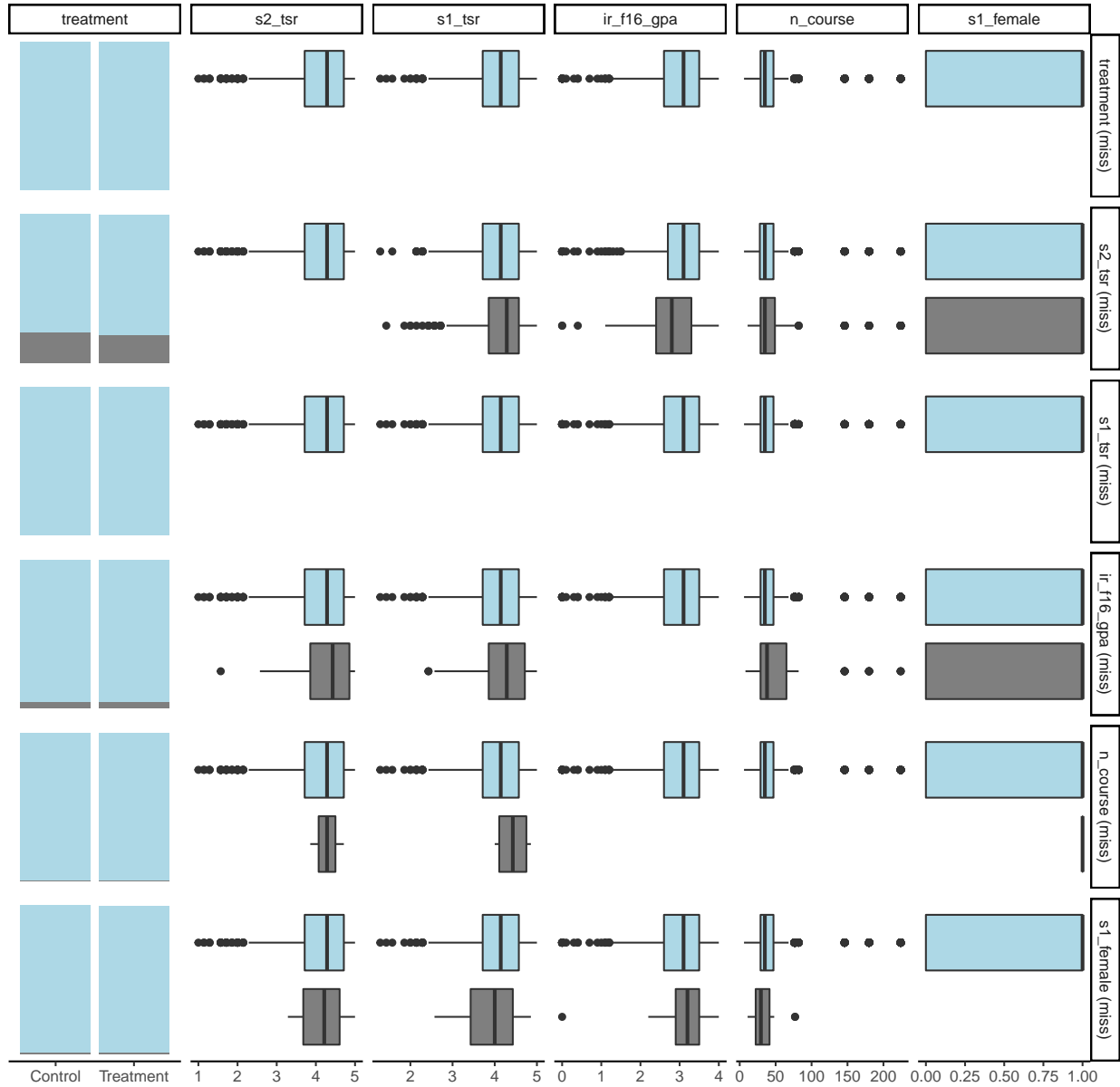
References

Abrami, Philip C., and Deborah A. Mizener. 1985. "Student/Instructor Attitude Similarity, Student Ratings, and Course Performance." *Journal of Educational Psychology* 77 (6): 693–702. https://doi.org/10.1037/0022-0663.77.6.693.

Bers, Trudy, and Pam Schuetz. 2014. "Nearbies: A Missing Piece of the College Completion Conundrum." *Community College Review* 42 (3): 167–83. https://doi.org/10.1177/0091552114525834.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2009. *College: What It Was, Is, and Should Be.* Princeton, U.S.A.: Princeton University Press.

Carly D. Robinson, Harvard University, Whitney Scott, and Michael A. Gottfried. 2020. "Instructor-Student Relationship Experiment: Data Package." *Inter-University Consortium for Political and Social Research [Distributor].* Inter-university Consortium for Political; Social Research (ICPSR). https://doi.org/10.3886/E119071V2.

Coolican, Hugh. 2014. *Research Methods and Statistics in Psychology.* 6th ed. Psychology Press. https://doi.org/10.4324/9780203769836.

Delbanco, Andrew. 2015. *Crossing the Finish Line: Completing College at America's Public Universities.* Princeton, U.S.A.: Princeton University Press.

Engineering, Stefan. 2018. "Create an Apa Style Correlation Table with R." *Blog Entry.* https://stefaneng.github.io/apa_correlation_table/; GitHub.

Evans, William N., Melissa S. Kearney, Brendan Perry, and James X. Sullivan. 2020. "Increasing Community College Completion Rates Among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case-Management Intervention." *Journal of Policy Analysis and Management* 39 (4): 930–65. https://doi.org/10.1002/pam.22256.

Gehlbach, Hunter, Maureen E. Brinkworth, and Anna D. Harris. 2012. "Changes in Teacher–Student Relationships." *British Journal of Educational Psychology* 82 (4): 690–704. https://doi.org/10.1111/j.2044-8279.2011.02058.x.

Gehlbach, Hunter, Maureen Elizabeth Brinkworth, Aaron King, Laura Hsu, Joe McIntyre, and Todd T. Rogers. 2015. "Creating Birds of Similar Feathers: Leveraging Similarity to Improve Teacher-Student Relationships and Academic Achievement." https://doi.org/10.1037/edu0000042.

Gilbert, Clark G., and Michael B. Horn. 2020. "A Certificate, Then a Degree." *Education Next* 20 (1): 86–87.

Gurantz, Oded, Michael Hurwitz, and Jonathan Smith. 2017. "BOOSTING Hispanic College Completion: Does High-School Recruiting Help More Students Graduate?" *Education Next* 17 (3): 60–67.

Lai, Tzu-Ling. 2015. "Effects of Student-Teacher Congruence on Students' Learning Performance: A Dyadic Approach." *Social Science Quarterly* 96 (5): 1424–35. https://doi.org/10.1111/ssqu.12215.

Robinson, Carly D., Whitney Scott, and Michael A. Gottfried. 2019. "Taking It to the Next Level: A Field Experiment to Improve Instructor-Student Relationships in College." *AERA Open* 5 (1): 1–15. https://doi.org/10.21105/joss.01686.